þÿ W a r V i c t i m S a m p o 1 9 1 4 1 9 2 2 : A S e m a n t i c P o r t a l a Data Service for Digital Humanities Research on War History

Rantala, Heikki

Springer
2020

Rantala , H , Ikkala , E , Jokipii , I , Koho , M , Tuominen , J & Hyvönen , E 2020 , þÿ W a r V i c t i m S a m p o 1 9 1 4 1 9 2 2 : A S e m a n t i c P o r t a l a n d L i n k e d D a t a S e r v i c Humanities Research on War History . in A Harth , V Presutti , R Troncy , M Acosta , A Polleres , J D Fernández , J Xavier Parreira , O Hartig , K Hose & M Cochez (eds) , The Semantic Web: ESWC 2020 Satellite Events . Lecture Notes in Computer Science , vol. 12124 , Springer , Cham , pp. 191-196 , Extended Semantic Web Conference 2020 , 02/06/2020 . https://doi.org/10.1007/978-3-030-62327-2_33

http://hdl.handle.net/10138/326062
https://doi.org/10.1007/978-3-030-62327-2_33

unspecified
acceptedVersion

WarVictimSampo 1914–1922: A Semantic Portal and Linked Data Service for Digital Humanities Research on War History

Heikki Rantala[1][0000−0002−4716−6564], Esko Ikkala[1][0000−0002−9571−7260], Ilkka Jokipii[2,3], Mikko Koho[1][0000−0002−7373−9338], Jouni Tuominen[1,2][0000−0003−4789−5676], and Eero Hyvönen[1,2][0000−0003−1695−5840]

[1] Aalto University, Semantic Computing Research Group (SeCo), Finland
https://seco.cs.aalto.fi, firstname.lastname@aalto.fi
[2] University of Helsinki, Helsinki Centre for Digital Humanities (HELDIG), Finland
https://heldig.fi, firstname.lastname@helsinki.fi
[3] The National Archives of Finland
https://arkisto.fi

Abstract. This paper presents the semantic portal and Linked Open Data (LOD) service WarVictimSampo 1914–22 about the war victims, battles, and prisoner camps in the Finnish Civil and other wars. The system is based on a database of the National Archives of Finland and related data compiled during the project. The system contains detailed information about some 40 000 deaths extracted from several data sources, and data about prisoner camps and over 1000 battles of the Civil War. A key novelty of WarVictimSampo 1914–22 is the integration of ready-to-use Digital Humanities tooling with the data service, which allows, e.g., studying information about wider prosopographical groups in addition to individual victims. We demonstrate how the tools of the portal, as well as the underlying SPARQL endpoint, can be used to explore and analyze war history in flexible and visual ways. WarVictimSampo 1914–22 is a new member in the series of "Sampo" model based semantic portals. It was published in late 2019 and got 20 000 users in two weeks.

Keywords: Linked Data · Semantic Web · War History.

1 Introduction

This paper presents the semantic portal and Linked Open Data (LOD) service WarVictimSampo 1914–22[4] about the war victims, battles, and prisoner camps in the Finnish Civil and other wars. The tools offered by the service help researchers and general public to better access the historical data. The main focus of the service is the Finnish War Victims 1914–1922 database that includes some 40 000 victims and is maintained by the National Archives of Finland.

[4] WarVictimSampo 1914–1922 Semantic Portal: https://sotasurmat.narc.fi/en

Most, over 90 percent, of the deaths recorded here are related to the Finnish Civil War in 1918, and the rest are related to the other wars of the period. Fig. 1 depicts the distribution of death dates in the data during 1918 as shown by the service. The data includes people who have died in Finland and abroad.

In this paper we show how LOD and modern web technologies can be used to enhance and update an old data service. The paper also demonstrates how modern tools can be used with LOD to analyze the data for Digital Humanities research. The original War Victims data was recorded in 1999–2003 as a government project [11] and includes 39 931 deaths. The original data was converted into LOD form and updated with 1590 new previously unknown victims and some new information concerning the old records [9]. An important contribution of the new system is making the access to the data more versatile and easier.

The old database is not directly open for public access. There is an old web application[5] in use for exploring the data with simple search functionality and a homepage for each person. A person's homepage includes basic information about the victim, but many pieces of information are not shown, even though they would be available in the underlying database. The end users of the system have deemed the search interface fairly inflexible and with too few options to choose from. Also means of exporting the data from the database ware asked for.

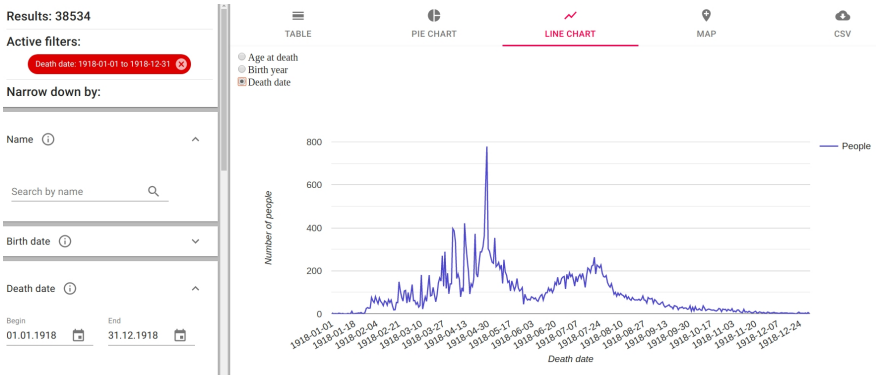

Fig. 1. Distribution of death dates in 1918 as shown in WarVictimSampo 1914–22

The death records contain basic information of the people (e.g., name, place of birth, date of birth, date of death), socioeconomic information (e.g., occupation, marital status), and war related information (e.g., military rank, military organization, time of imprisonment). During the conversion to LOD form the data was ontologized where appropriate, and mapped to some outside sources. This made it easy to, for example, create links to outside web pages relating to

[5] http://vesta.narc.fi/cgi-bin/db2www/sotasurmaetusivu/main?lang=en

certain famous people in the data, and to get coordinates for the municipalities to use with map visualisations.

To publish the data we use the "Sampo" publishing model [4]. The data in LOD form is loaded into a triplestore hosted at the Linked Data Finland platform[6] [3], where it can be queried using SPARQL. The semantic portal makes queries to this publicly open endpoint, and a researcher can also query the endpoint for her own purpose.

2 WarVictimSampo 1914–22 Semantic Portal

A semantic portal was developed to allow different user groups to access the data easily. The user groups include researchers, students, and the wider public interested in either the Finnish Civil War in general or the fates of their relatives. Even though the data can be accessed by anyone with SPARQL queries this can be too technically demanding for many users. Also a researcher who is able to create her own SPARQL queries finds it useful to have an easy way to explore the data and to create simple visualizations quickly. The visualization tools provided by the portal are expected to be useful for both finding new data and for educating the public about history. These tools should not be expected to fully replace manual research and close reading. They are aimed to be used to spot interesting phenomena in the data that require more detailed analysis.

The user interface of the semantic portal is implemented as a full stack JavaScript web application, using the Sampo-UI framework[7] [6]. The user interface is built around the concept of faceted search [10]. With faceted search, the user can easily narrow the search step by step by making selections based on predetermined orthogonal hierarchies of property values called facets. Facets also show the number of available items with each possible selection. This allows the user to immediately see the number of solutions of each possible selection. Combined with selections on other facets like occupation, party, and age, the user may also draw interesting conclusions by observing the hit distributions on the facets. Faceted search can therefore be used to not only find individuals that fit certain criteria, such as relatives, but it can also be used to find information about the distributions of different kind of the casualties. The faceted search paradigm is an example of exploratory search [7].

The user interface currently includes two main perspectives for exploring the underlying knowledge graph: 1) The main perspective is based on searching and exploring the casualties. 2) There is also a perspective based on the battles of the Finnish Civil War, covering currently 1182 geo-coded battles. Other views may be added later in the same way as in other "Sampo" series semantic portals[8].

For the both perspectives there are multiple tabs to view the data in different ways. Currently the data can be shown as a table or downloaded as a CSV

[6] http://www.ldf.fi/dataset/siso

[7] https://github.com/SemanticComputing/sampo-ui

[8] https://www.europenowjournal.org/2019/09/09/linked-data-in-use-sampo-portals-on-the-semantic-web

file, and it can be visualized with pie charts, line charts, and maps. All the visualizations are dynamic, reflecting the selections made with the facets. This allows for visualizing different aspects of the data quickly and easily. For example, Fig. 2 shows how people in the data from the Vaasa province clearly have died mainly in the city of Tampere for some reason that can be explored further.

The table view can be considered the default view and can be used to browse the individual war victims in the data. This table shows the names of the people and their basic information. By clicking the name of the person, the user can navigate to that person's homepage in the application. The person's page shows all the information related to the person, including the sources of all the individual pieces of information.

Both the line chart and pie chart visualizations can be used for multiple different purposes. For example, the line chart can be used to visualize age distributions, birth years, or death dates as shown in Fig. 1. The line chart can be used, for example, by researchers to compare the age distributions of different groups of people within the data by using one or more facets to select the desired group of people. Line charts could also be used, for example, by school students to see the dates when people from their hometown died during the Finnish Civil War. The pie chart can be used to visualize distributions within one facet. For example, it could be used to visualize the relative amount of people with different occupations in the data.

The Battles perspective includes an animation view that can be used to visualize the battles of the Finnish Civil War. The animation advances by day and shows the battles that start that day as red dots that stay on the map, but turn dark when the time advances. The purpose of this view is mainly educational.

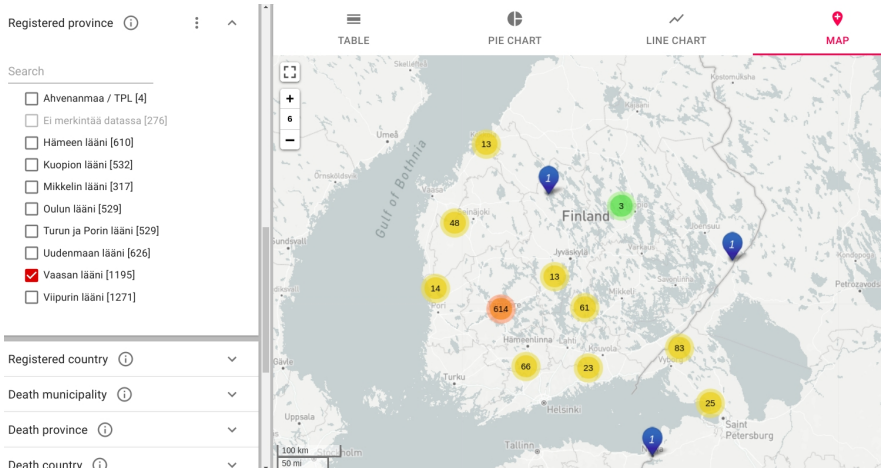

Fig. 2. A map visualization of death places of the people who were registered in Vaasa province and supported Whites in the Finnish Civil War.

The WarVictimSampo 1914–22 semantic portal and data service were opened to public on 20 November, 2019, and has had tens of thousands of users. There is an English translation of the user interface available, but all the data are only in Finnish.

The plan is to continue developing the portal and updating the data service. For example, a perspective of the prison camps of the Finnish Civil War could be added to the portal. Because the data and the source code of the user interface of the semantic portal are open to all, new views and perspectives could be developed by anyone interested. The user interface of the portal could also be used as a model for creating user interfaces for some other data services.

3 Related Work

WarVictimSampo 1914–22 is a follow up project of WarSampo [5], which uses LD to present and publish information related to the Second World War in Finland, including death records. The novelty of WarVictimSampo 1914–22 lays in the idea of developing new data-analytic tooling for research in war history, as well as in creating, cleaning, extending, and publishing the former War Victims 1914–22 database for open use on the the Semantic Web.

There have been several projects publishing linked data about the World War I on the Web, such as Europeana Collections 1914–1918[9], 1914–1918 Online[10], WW1 Discovery[11], Out of the Trenches[12], CENDARI[13], Muninn[14], and WW1LOD [8]. In addition to WarSampo, there are a few works that use the Linked Data approach to WW2, such as [2,1], Open Memory Project[15] on holocaust victims, and the Dutch project Netwerk Orloogsbronnen[16].

Acknowledgements Our research was funded by the Prime Minister's Office. Thanks to CSC – IT Center for Science, Finland, for computational resources.

References

1. de Boer, V., van Doornik, J., Buitinck, L., Marx, M., Veken, T.: Linking the kingdom: enriched access to a historiographical text. In: Proceedings of the 7th International Conference on Knowledge Capture (KCAP 2013). pp. 17–24. ACM (2013)

[9] http://www.europeana-collections-1914-1918.eu
[10] http://www.1914-1918-online.net
[11] http://ww1.discovery.ac.uk
[12] http://www.canadiana.ca/en/pcdhn-lod/
[13] http://www.cendari.eu/research/first-world-war-studies/
[14] http://blog.muninn-project.org
[15] http://www.bygle.net/wp-content/uploads/2015/04/Open-Memory-Project_3-1.pdf
[16] https://www.oorlogsbronnen.nl

2. Collins, T., Mulholland, P., Zdrahal, Z.: Semantic Browsing of Digital Collections. In: Proceedings of the 4th International Semantic Web Conference (ISWC 2005). pp. 127–141. Springer–Verlag (2005)
3. Hyvönen, E., Tuominen, J., Alonen, M., Mäkelä, E.: Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. In: The Semantic Web: ESWC 2014 Satellite Events. pp. 226–230. Springer–Verlag (2014)
4. Hyvönen, E.: "Sampo" Model and Semantic Portals for Digital Humanities on the Semantic Web. In: Proceedings of the Digital Humanities in the Nordic Countries (DHN 2020). CEUR Workshop Proceedings (2020), accepted
5. Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., Mäkelä, E.: WarSampo Data Service and Semantic Portal for Publishing Linked Open Data about the Second World War History. In: The Semantic Web – Latest Advances and New Domains (ESWC 2016). pp. 758–773. Springer–Verlag (2016)
6. Ikkala, E., Hyvönen, E., Rantala, H., Koho, M.: Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces (2020), `https://seco.cs.aalto.fi/publications/2020/ikkala-et-al-sampo-ui-2020.pdf`, submitted
7. Marchionini, G.: Exploratory search: from finding to understanding. Communications of the ACM **49**(4), 41–46 (2006)
8. Mäkelä, E., Törnroos, J., Lindquist, T., Hyvönen, E.: WW1LOD: An application of CIDOC-CRM to World War 1 Linked Data. International Journal on Digital Libraries **18**(4), 333–343 (2017)
9. Rantala, H., Jokipii, I., Koho, M., Ikkala, E., Tuominen, J., Hyvönen, E.: Building a Linked Open Data portal of war victims in Finland 1914–1922. In: Proceedings of Digital Humanities in Nordic Countries (DHN 2020). CEUR Workshop Proceedings (2020), accepted
10. Tunkelang, D.: Faceted search. Synthesis Lectures on Information Concepts, Retrieval, and Services, Morgan & Claypool, Palo Alto, California (2009)
11. Westerlund, L. (ed.): Sotaoloissa vuosina 1914—22 surmansa saaneet. Tilastoraportti. Valtioneuvoston kanslia, Helsinki, Finland (2004)